This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problems Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 17/30 — A1

(11) International Publication Number: WO 97/38378

(43) International Publication Date: 16 October 1997 (16.10.97)

(21) International Application Number: PCT/US97/05782

(22) International Filing Date: 8 April 1997 (08.04.97)

(30) Priority Data:

60/017,912	10 April 1996 (10.04.96)	US
08/826,940	8 April 1997 (08.04.97)	US

(71) Applicant: AT & T CORP. [US/US]; 32 Avenue of the Americas, New York, NY 10013-2412 (US).

(72) Inventor: KIRK, Thomas; 22 King George Road, Warren, NJ 07059 (US).

(74) Agent: DWORETSKY, Samuel, H.; AT & T Corp., P.O. Box 4110, Middletown, NJ 07748-4110 (US).

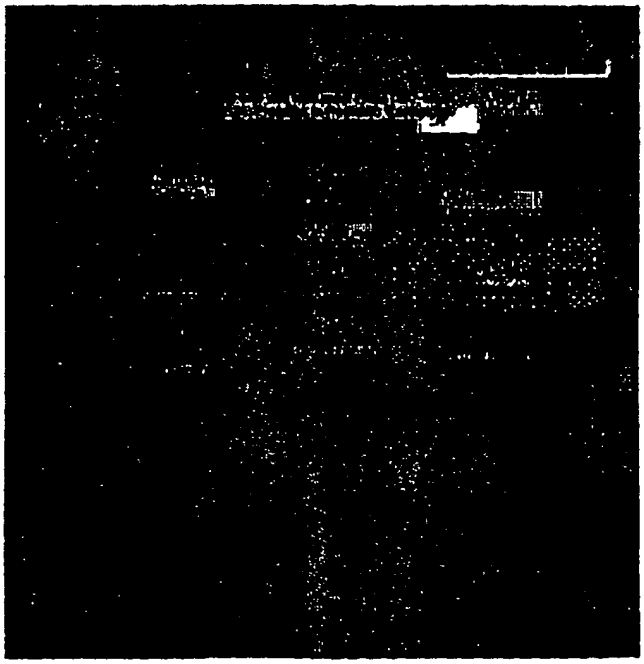
(81) Designated States: CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD OF ORGANIZING INFORMATION RETRIEVED FROM THE INTERNET USING KNOWLEDGE BASED REPRESENTATION

(57) Abstract

A system and method of organizing electronic representation of documents in a knowledge based representation system is disclosed. The knowledge based representation system operates in an environment where computers and networks are interconnected and where documents can be retrieved from the computers and networks. A query is created to search for the documents. The system determines which of the computers and networks are capable of understanding the query syntax. The query is sent to each of the computers and networks that can handle the query. The system receives results relating to the documents from the computers and networks. The results are merged into a single result set. Each of the results contains a reference to each of the documents. The documents are then refined by comparing the documents with text matching patterns of the knowledge base. Refining is accomplished by retrieving a document for each of the references and then applying the matching patterns to the documents. The system determines a list of concepts that match the documents. The system provides the documents to the knowledge based representation system as instances of the concepts.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

METHOD OF ORGANIZING INFORMATION RETRIEVED FROM THE INTERNET USING KNOWLEDGE BASED REPRESENTATION

Technical Field

This invention relates to the field of accessing information on the Internet and, more particularly, to a method of organizing information retrieved from the Internet using a knowledge based representation system.

Background of the Invention

The Internet is a series of inter-connected networks which facilitate the exchange of information, data, and files. Users connected to the Internet have access to the vast amount of information on these networks. A typical way of getting access to the Internet is through an online service server. Referring to FIG. 1, networks 110, 112, and 114 are connected to Internet 100 via online service servers 120, 122, and 124, respectively. Another way of getting access to the Internet is through a dial-in Internet provider. For example, a user on his personal computer ("P.C.") 158 may access Internet 100 by dialing in to Internet provider 150 using his modem 152. Routers, which connect computers and networks, direct traffic in a network and on the Internet. Routers 160, 162, 164, and 166 examine packets of data that travel across the networks and Internet to determine where the data is headed.

Online service servers and Internet providers allow users to search the World Wide Web ("Web"), a globally connected network on the Internet, using software programs known as search engines 130, 132, 134, and 154. Search engines are also known as search tools and Web crawlers. These search engines travel across the Web gathering documents by following the hypertext links found in Web (home) pages 140, 142, 144, and 156.

One way of searching the Internet is by keywords. For example, a user types in a query string of keywords that describes the information he is looking for. The search engine searches databases on the Internet and results are returned in hypertext markup language ("HTML") pages. A user can then view a document of interest by "clicking" on a link to that document. Clicking refers to the process of actuating a mouse switch by centering a cursor on the desired item.

While present search engines provide for searching of keywords on the Internet, the vast amounts of information on the Internet makes getting relevant information difficult. Stated another way, keyword searches typically result in a return of vast amounts of information that the user must browse through in order to retrieve the relevant information. Thus, what is required is a more effective method of retrieving information from the Internet.

Summary of the Invention

The above-stated problem of organizing information search results is mitigated by the application of knowledge based representation techniques for automatically categorizing search results. This information retrieval and management system associates a knowledge base with search servers to improve the relevance and precision of search tasks. The knowledge base provides a user profile (topic taxonomy) that reflects the interests and preferences of the user for organizing information. The system uses this knowledge base to organize the results of keyword searches. The system automatically categorizes and segments search results in accordance with the knowledge base to provide for easy searching of relevant information. The system displays the search results over a subset of the knowledge based topic taxonomy, segmenting the results in a way that makes it easy to find the most relevant documents, and filtering out irrelevant results.

Brief Description of the Drawings

In the drawing,

FIG. 1 illustrates a diagram of computers and networks and their connection to the Internet for discussion of the environment in which the present invention operates;

FIG. 2 is a block diagram of an exemplary knowledge based browser displaying a graphical representation of a concept generalization taxonomy in accordance with the principles of the present invention;

FIG. 2a is an actual screen display of the exemplary knowledge based browser of FIG. 2;

FIG. 3 is a block diagram of a search interface in accordance with the principles of the present invention; and

FIG. 4 shows a flow diagram illustrating the steps required for a user to retrieve information from the Internet and organize it using knowledge based representation.

Detailed Description

Referring to FIG. 1, there is shown an environment for the present invention including exemplary networks 110, 112, and 114 and P.C.'s 158 and 159 which are inter-connected to Internet 100. These networks comprise users who are connected to one another in, for example, a token ring network (network 114) or through an Ethernet network (networks 110 and 112). Each network further comprises a server 120, 122, and 124. A server is a host computer that allows users to communicate with each other on the network or with users outside the network through the Internet. Users on P.C.'s 158 and 159 may subscribe to Internet Provider 150, which allows users to communicate with each other and other users on the Internet.

Any user may search for information available on the Internet. If a computer or network is connected to the Internet, then information on that computer or network is accessible by others if it is not protected. Since the Internet is a global network, the amount of information that can be retrieved is immense. Many servers and providers include search engines 130, 132, 134, and 154 that allow users to search by keywords. These search engines are computer programs which are search-application based programs that run on online service servers 120, 122, 124, and Internet provider 150. Searching by keywords typically results in a return of vast amounts of information that the user must browse through in order to get the desired information.

Currently, there are two ways of searching the Internet. Both methods operate under a client/server model. By client/server model is intended a user running a piece of software on his computer or a shared program of a server--the client--to use the resources of a distant server computer (other servers connected on the Internet). For example, in FIG. 1, a user on P.C. 110a may search for information on online service servers 122 and 124 and Internet provider 150. Similarly, a user on P.C. 156 may search for information on online service servers 120, 122, and 124. The distant servers, e.g., online service servers 120, 122, and 124 and Internet provider 150, are also called hosts because they serve many users of many networks. The hosts allow many different clients to access their resources at the same time; the hosts are not devoted to a single user.

The first way of searching the Internet is through indexes. Indexes present a highly

structured way of finding information. Indexes let users browse through information by categories such as arts, computers, entertainment, sports, etc. In a Web browser, a user on his P.C. 110a can click on a category by, typically, using his mouse 110b and is presented with a series of subcategories. For example, under sports a user may find baseball, basketball, football, etc. Depending on the size of the index, there may be several layers of subcategories. When the user gets to the subcategory he is interested in, he will be presented with a list of relevant documents. To get to those documents, the user clicks on links to them. "Yahoo!" is the name of a popular index on the Internet. Yahoo! and other indexes also allow users to search through them by typing in words that describe information that the user is looking for. The user then gets a set of search results--links to documents that match his search. To get the information, the user clicks on a link to the document.

The second way of finding information is to use search engines, also known as search tools. Search engines operate on essentially static pre-built indexes, i.e., the indexes are built up from online content and stored in a database on a search server. Web crawlers are used by the search engines for gathering the online content that is retrieved and indexed in the search server's database. Some popular Internet search engines include Lycos, WebCrawler, and Alta Vista. To begin a search, a user types in keywords that describe the information he wants. Results that match the user's search criteria from the search are sent back to the user. From the list of results, the user can retrieve a document by clicking on a link to that document.

Although both indexes and search engines allow users to find information on the Internet, the information found is typically large and often difficult to locate relevant information. Therefore, it is desirable to automatically categorize search results found on the Internet so as to allow users to easily browse through the search results to find relevant information.

According to the present invention, knowledge based representation systems, with their capabilities for representing and inferring relationships among objects, mitigate the above problems. In particular, the present invention is directed to a knowledge based information retrieval and management system that enhances searches on any multi-network system such as the Internet. The system provides users with means to superimpose a tailored conceptual

organization over the information found on the Internet, thereby enriching the usefulness of and access to that information. Referring to FIG. 1, the system is integrated with existing Web browsers 130, 132, 134, and 154 to create a seamless environment combining hypertext browsing with conceptual navigation. The system may also be stored on a personal computer, e.g., P.C. 110a, in which case only users with access to that personal computer may use the system.

Referring now to FIG. 2, it illustrates an exemplary knowledge based browser which displays a graphical representation of a concept generalization taxonomy 200 in accordance with the present invention. A taxonomy is a generalization hierarchy which graphically displays relationships between concepts. A concept is an abstract description of an object. Nodes in FIG. 2 correspond to knowledge base concepts (e.g., 210, 220, 230, 212, 214, etc.), and edges (e.g., 210a, 210b, 220a, etc.) connecting the nodes indicate subsumption relationships between the concepts. A feature of the present invention is the system can manage the subsumption relationships automatically based on concepts and instances (270, 280). An instance is a specific realization of a concept, i.e., a concept is an abstract description of something while an instance of that concept is a real object that satisfies that description. For example, when a new document is added to the knowledge based browser as an instance, the system infers all the places it belongs in the taxonomy.

As illustrated in FIG. 2, the most general concepts are at the left. Following outgoing edges of a concept node (going from left to right) leads to more specialized concepts. For example, the topic "artificial intelligence" 228 is a specialization of "computer science" 220, and "knowledge representation" 229 is in turn a specialization of "artificial intelligence" 228. The panels 270 and 280 within this display show lists of instances of these concepts. For example, the panel 270 shows documents which are instances of the topic "pediatric medicine" 212; the panel 280 shows instances of the concept "knowledge representation" 229. Instances are inherited by parent concepts all the way up the hierarchy, so for example, the documents appearing under "knowledge representation" would also appear under "computer science". The method of organizing instances is discussed below with regard to a search interface. FIG. 2a is an actual screen display of the exemplary knowledge based browser of FIG. 2, illustrating the

concept generalization taxonomy 200 and the subsumption relationships between concepts and instances.

The search interface operates similarly to that of the knowledge base browser. The search interface uses a knowledge base to refine search results by segmenting and categorizing results with respect to a user's concept generalization taxonomy. For example, after results from a keyword search have been combined in a result set for display, the system provides an additional refining step that can further focus the result set. Refining the result set against the knowledge base involves retrieving the documents in the result set and processing them with the knowledge base pattern matchers. Textual patterns associated with concepts in the knowledge base allow the knowledge representation system to categorize and organize these documents within the concept taxonomy. Each pattern in the knowledge base is associated with a concept. Stated another way, each document is compared against these pattern matchers to determine whether there are any concepts that match the document. The output of this comparison process is a set of specific concepts in the knowledge base that have some correspondence to the content of the document. A record of a match between a concept and the document is made in the knowledge base by creating a temporary instance whose description includes the matched concepts. Finally, the refined search result is presented graphically over a subset of the knowledge base topic taxonomy. This subset is defined by those concepts having one or more of the temporary instances created during the matching process. This is illustrated in FIG. 3 where only those concepts that match the contents of a document are displayed.

The present invention of using a knowledge based representation system in organizing data is especially helpful when a keyword search results in thousands of documents. By running pattern matchers against those documents, one can quickly narrow down those documents that are most relevant to the user.

Accordingly, the knowledge based representation system (browser and search interface) of the present invention allow users to quickly find relevant information.

Another feature of the taxonomy is that by grouping the results according to concepts, a user may zoom in on the part that he thinks is most relevant. This further enhances searching on

the Internet by saving browsing time.

The search interface further implements transparent, concurrent access to multiple index servers in order to maximize query coverage and minimize response latency. By explicitly representing the capabilities of the individual search engines, the query system ensures that only those index servers capable of handling the query are consulted.

Another feature of the present invention is a user interface which provides editors for extending and reorganizing the concept hierarchy. The user interface also provides for a navigation browser that maintains an interactive graphical map of the navigation history. The navigation browser is a tree-structured graphical representation of the user's browsing history. Its function is as follows: as the user browses, he generates an ordered sequence of the web sites he visits, following links from one page to another. As he backtracks and makes new browsing choices, the browsing history becomes a branching tree. The navigation browser keeps track of these choices adding new nodes to the tree for every site/page visited. This tree, besides showing an overview of the browsing history, becomes an alternative way to navigate (by clicking on the node in the tree to return to the associated page).

Another feature of the present invention is that the system architecture separates the knowledge base from the client to allow the user to maintain a consistent view of his information space regardless of the client's location. By keeping the knowledge base in one place, the environment can follow the user from one platform to another. An advantage of the separation is to help ensure continuous availability of the system server since it provides shared access to the knowledge base and performs autonomous monitoring tasks even when the client is inactive or disconnected. In other words, the knowledge base may be stored on another server, separated from the client.

Referring to the flowchart in FIG. 4, this flowchart illustrates the steps required for a user to retrieve information from the Internet and organize it using a knowledge based representation system in accordance to the present invention.

In step 401, a user enters a query string of keywords to be searched on his personal computer 110a using a knowledge based Web browser 130 in accordance with the present

invention. The knowledge based Web browser is a software that may be installed in either a client 110a or server 120.

In step 403, the query string is pre-processed to determine which search servers are capable of understanding the query syntax. This is done by examining the Universal Resource Locator ("URL") of the query string to determine which server(s) to send the request to. Generally, the query has to be translated into specific query syntax of the server that the user is requesting information. Typically, a query translator is provided with an interface to the server for serving the query.

In step 405, queries are sent to each server that can handle the expression. Queries may be sent out serially or concurrently. An advantage of sending out the queries concurrently is reduction of latency in both the network and search process. In other words, all servers can work on a query at the same time.

In step 407, depending on the result size threshold, individual servers may need to be queried repeatedly in order to gather the specified number of matches. Most servers, in order to limit the amount of resources that are used for a given query, will break the results coming back into some reasonable sets that are returned. For example, if there is a hundred hits for a search, a server may be set up to return only ten hits at a time. As such, if the specified number of matches is reached, then the procedure proceeds. If the specified number of matches has not been reached, then the servers are repeatedly queried until it has been reached.

In step 409, the results that come back from the servers are merged into a single result set. The results are merged by removing duplicates of the results. Each item in the result set consists of a reference to a document (a URL) and possibly a single line of descriptive text.

In step 411, if the user desires further refinement of the result set, he can request that the results be compared against the knowledge base pattern matchers. Else, the result set is displayed for the user.

In step 413, the document for each reference in the result set is retrieved.

In step 415, the pattern matcher(s) is applied to the document text to determine whether there are any topic concepts that match the text.

In step 417, a list of topic concepts that match the text of the document are generated.

In step 419, an instance is created for each document that matches a concept.

In step 421, the instance for the document is classified in the knowledge base's topic taxonomy.

The above iteration, steps 413-421, is parallelized to minimize the effects of network latency in gathering the text, since the result set may contain dozens or hundreds of documents to retrieve.

As the documents are retrieved and classified, the system incrementally displays the post-processed results graphically over a subset of the topic taxonomy, where the subset is defined by the collection of concepts having one or more instances from the search result. This is done to categorize and segment the search result with respect to concepts that are familiar and meaningful to the user. As such, by using the knowledge based representation system of the present invention, the search result may be browsed at various levels of detail, depending on how specific one wishes the segments to be.

What has been described is merely illustrative of the application of the principles of the present invention. Other arrangements and methods can be implemented by those skilled in the art without departing from the spirit and scope of the present invention.

I claim:

1. An apparatus for classifying electronic representation of documents, the apparatus comprising:

a knowledge based representation system which automatically organizes concepts and instances of concepts,

means for associating each of the concepts with a search pattern, and

means for using search patterns to determine whether each of the documents is an instance of the concepts and if so, providing document data to the knowledge based representation system as an instance of that concept.

2. The apparatus of claim 1 wherein:

the apparatus is employed in a system having a plurality of search engines and the apparatus further comprises

means for translating said search patterns into forms proper for the search engines and providing those forms to the search engines.

3. A method of organizing electronic representation of documents in a knowledge based representation system, the knowledge based representation system operates in an environment where computers and networks are interconnected and where the documents can be retrieved from the computers and networks, comprising the steps of:

entering a query to search for the documents,

determining which of the computers and networks are capable of understanding said query syntax,

sending said query to each of the computers and networks that can handle the query,

receiving results relating to the documents from the computers and networks,

merging said results into a single result set,

retrieving document data from said result set, and

refining said document data by comparing the documents with text matching patterns of said knowledge based representation system.

4. The method of claim 3, wherein each of said results consists of a reference to each of the documents.

5. The method of claim 4, wherein the refining step further comprising the steps of:

retrieving the documents for each of said references,

applying said matching patterns over the documents,

determining a list of concepts that match the documents, and

providing the documents to the knowledge based representation system as instances of said concepts.

6. The method of claim 3, wherein the sending step concurrently sends said query to each of the computers and networks that can handle the query.

7. The method of claim 5, wherein the retrieving step is done simultaneously for each of said references to minimize the effects of network latency in gathering said results.

8. The method of claim 5, wherein said text matching patterns may be edited by changing said list of concepts.

9. The method of claim 3, wherein the knowledge based representation system is stored in a client.

10. The method of claim 3, wherein the knowledge based representation system is stored in a server.

11. An apparatus for locating electronic representation of documents containing information about a given topic in an environment which includes means for using a matching pattern to locate documents, the apparatus comprising:

an information retrieval system wherein information is organized according to topics,

means in the information retrieval system for associating each of the topics with a matching pattern, and

means in the information retrieval system for responding to a query involving the given topic by providing the matching pattern associated with the given topic to the means for using a matching pattern to locate document data and returning at least the location of the document located by the means for using a matching pattern.

12. The apparatus of claim 11 wherein:

the environment includes a plurality of means for using a matching pattern to locate documents and the apparatus further comprises:

means for translating the matching pattern into forms proper to each of the means for using a matching pattern in the plurality thereof and providing the proper form to each of the means for using a matching pattern.

13. The apparatus of claim 11 further comprising:

means responsive to matching patterns for each of the given topics for using the matching patterns to determine which topics the document is an instance of and associating at least the location of the document with each of the given topics for which the matching patterns associated with the topics find matches in the document.

14. The apparatus set forth in claims 11, 12, or 13 further comprising

interactive receiving means in the means for associating which receive the matching pattern from a user of the apparatus.

15. The apparatus set forth in claim 11, 12, or 13 wherein

the matching pattern includes an expression in a regular expression language.

16. A computer-readable medium for classifying electronic representation of documents, comprising:

a knowledge based representation component for automatically organizing concepts and instances of concepts,

one or more search pattern components, for associating each of the concepts with a search pattern, and

a text matching component, for using search patterns to determine whether each of the documents is an instance of the concepts and if so, providing document data to the knowledge based representation system as an instance of that concept.

17. The computer-readable medium of claim 16 wherein:

the computer-readable medium is employed in a system having a plurality of search engines and the computer-readable medium further comprises

one or more translating components, for translating said search patterns into forms proper for the search engines and providing those forms to the search engines.

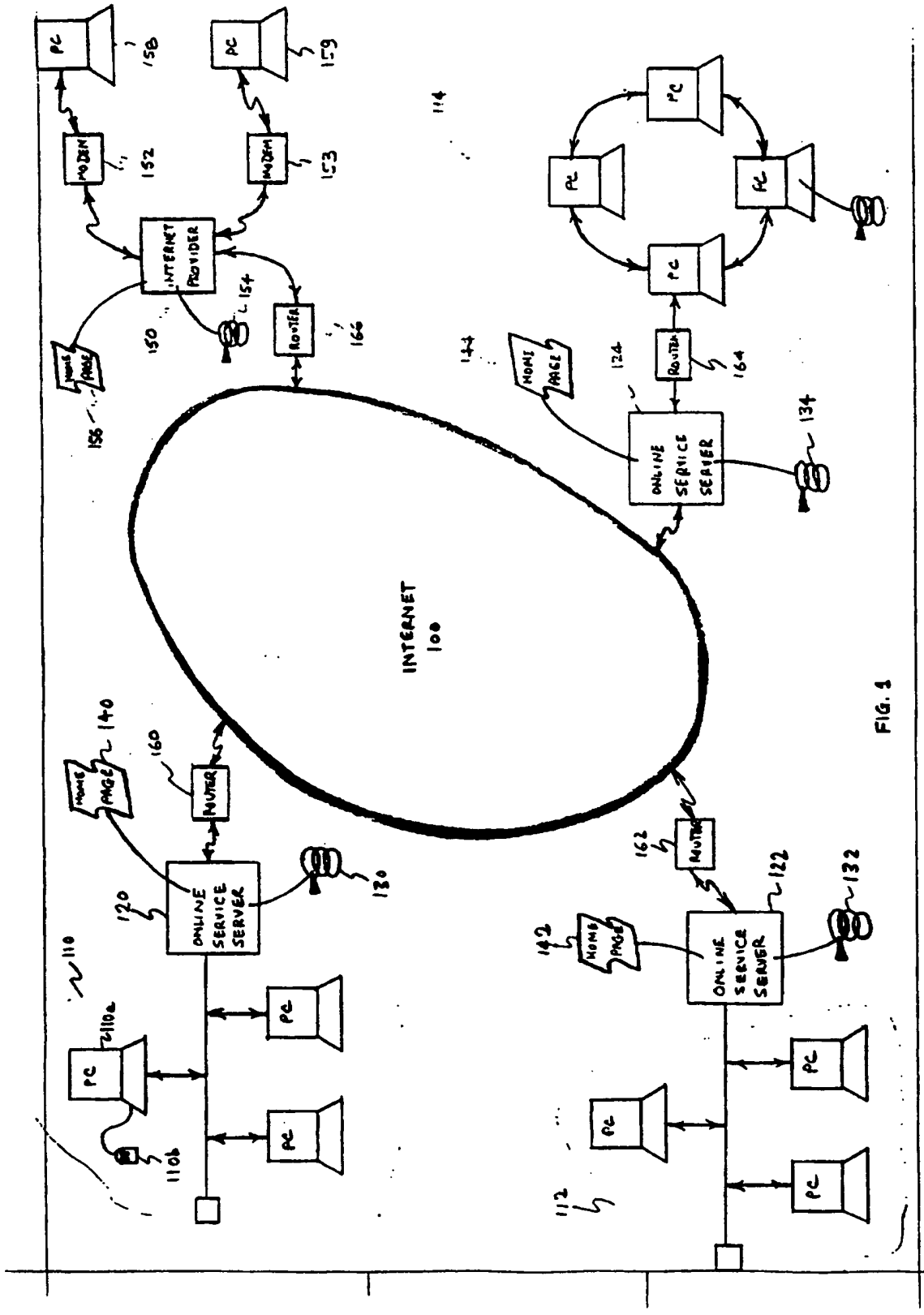

INTERNET
100
ONLINE SERVICE SERVER
120
130
HOME PAGE
140
ROUTER
160
110
110a
110b
112
ONLINE SERVICE SERVER
122
132
HOME PAGE
142
ROUTER
162
ONLINE SERVICE SERVER
134
HOME PAGE
144
ROUTER
164
124
INTERNET PROVIDER
150
MODEM
152
MODEM
153
154
ROUTER
166
HOME PAGE
156
PC
158
PC
159
114

FIG. 1

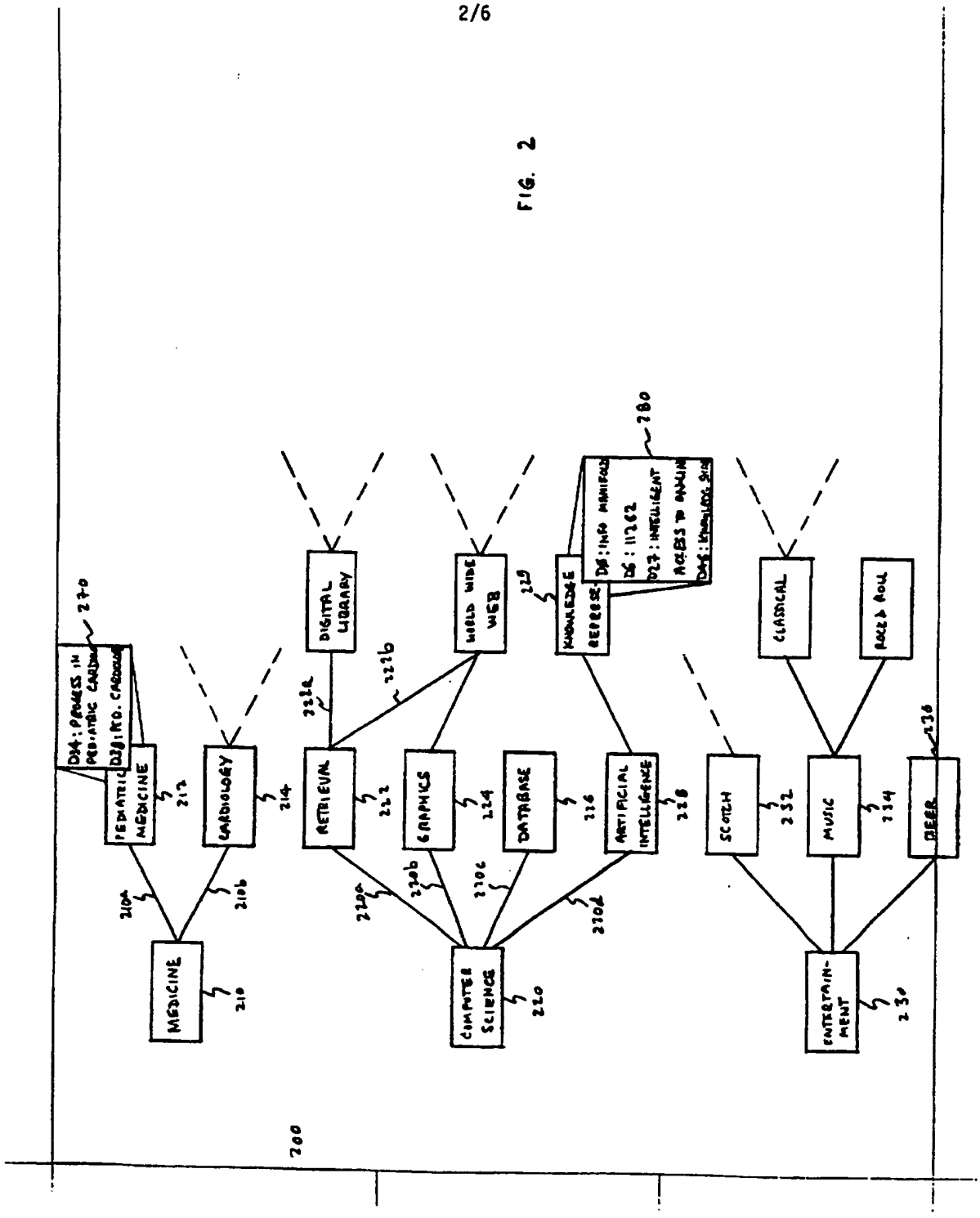

MEDICINE
210
PEDIATRIC MEDICINE
212
CARDIOLOGY
214
RETRIEVAL
222
GRAPHICS
224
DATABASE
226
ARTIFICIAL INTELLIGENCE
228
COMPUTER SCIENCE
220
ENTERTAIN-MENT
230
SCOTCH
232
MUSIC
234
BEER
236
CLASSICAL
ROCK & ROLL
DIGITAL LIBRARY
WORLD WIDE WEB
KNOWLEDGE REPRESE
280
210A
210B
220A
220B
220C
220D
222A
222B
224
200

FIG. 2

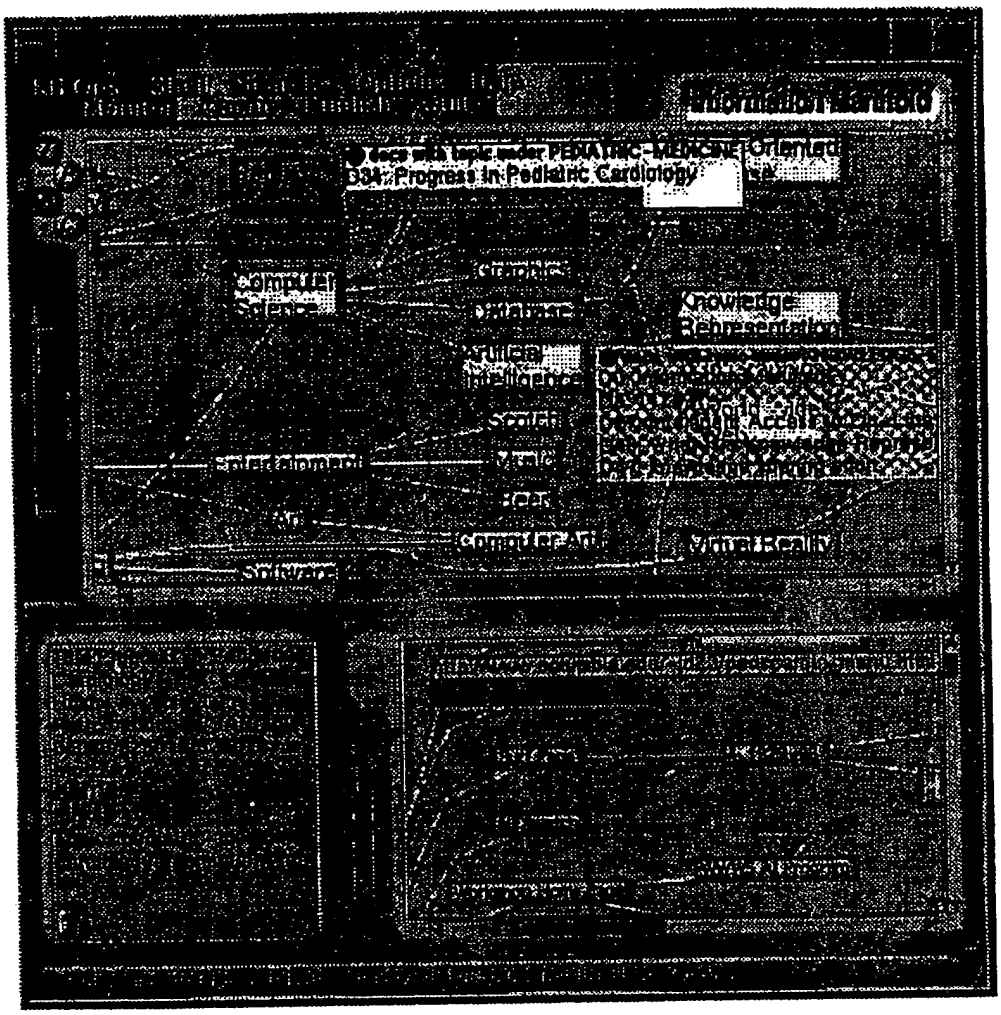

Information Manifold
Knowledge Representation
Virtual Reality

FIG. 2a

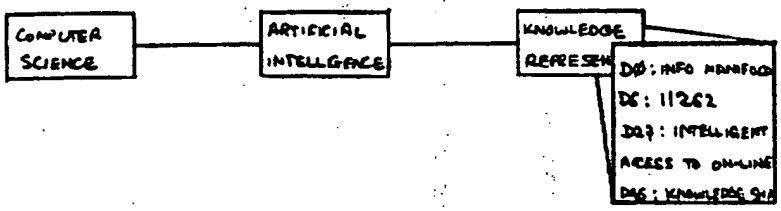
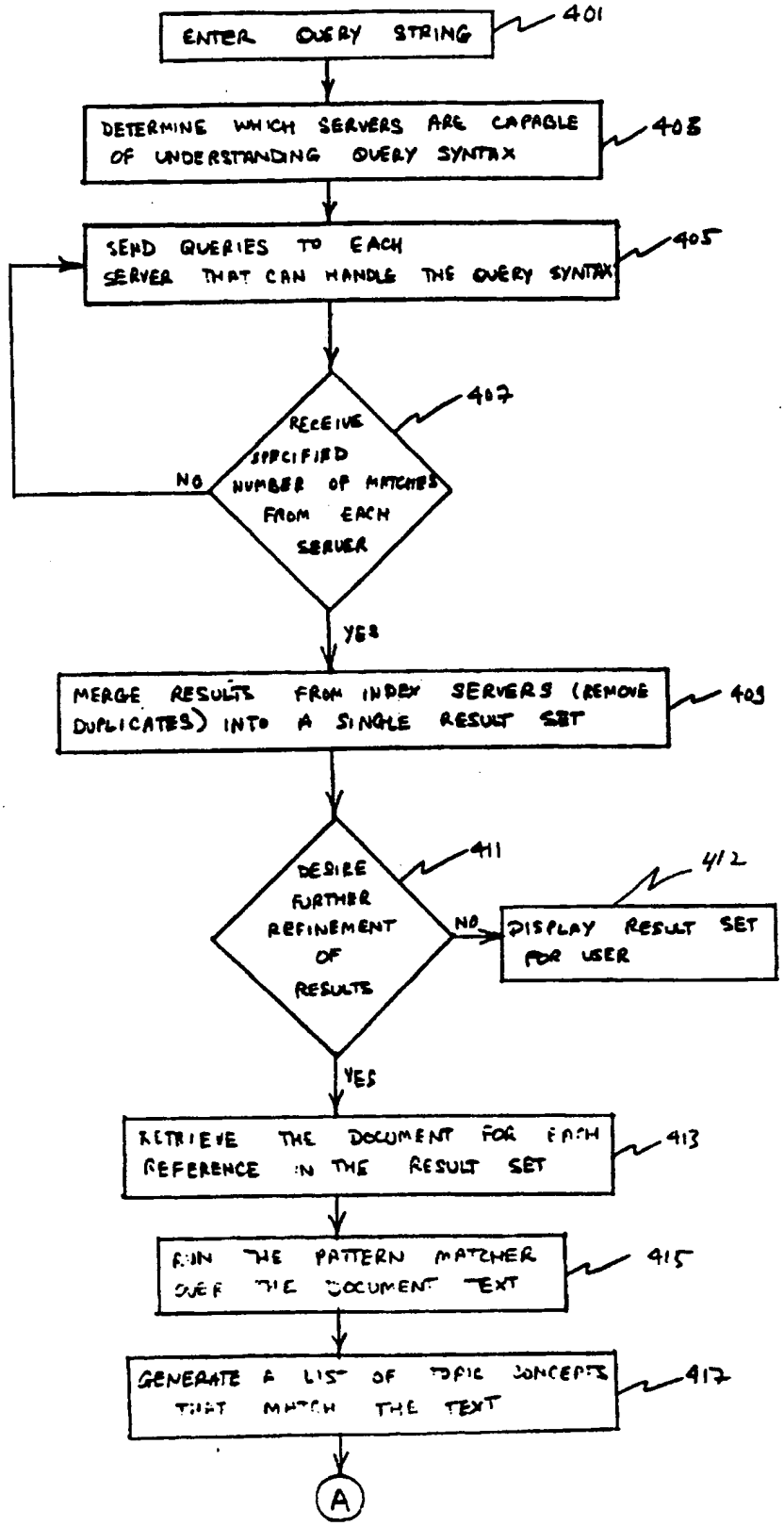

COMPUTER SCIENCE
ARTIFICIAL INTELLIGENCE
KNOWLEDGE
DØ: INFO MANIFOLD
D1: 11262
D2: INTELLIGENT
ACESS TO ON-LINE

FIG. 3


ENTER QUERY STRING
401
DETERMINE WHICH SERVERS ARE CAPABLE OF UNDERSTANDING QUERY SYNTAX
403
SEND QUERIES TO EACH SERVER THAT CAN HANDLE THE QUERY SYNTAX
405
RECEIVE SPECIFIED NUMBER OF MATCHES FROM EACH SERVER
407
NO
YES
MERGE RESULTS FROM INDEX SERVERS (REMOVE DUPLICATES) INTO A SINGLE RESULT SET
409
DESIRE FURTHER REFINEMENT OF RESULTS
411
NO
412
DISPLAY RESULT SET FOR USER
YES
RETRIEVE THE DOCUMENT FOR EACH REFERENCE IN THE RESULT SET
413
RUN THE PATTERN MATCHER OVER THE DOCUMENT TEXT
415
GENERATE A LIST OF TOPIC CONCEPTS THAT MATCH THE TEXT
417
A

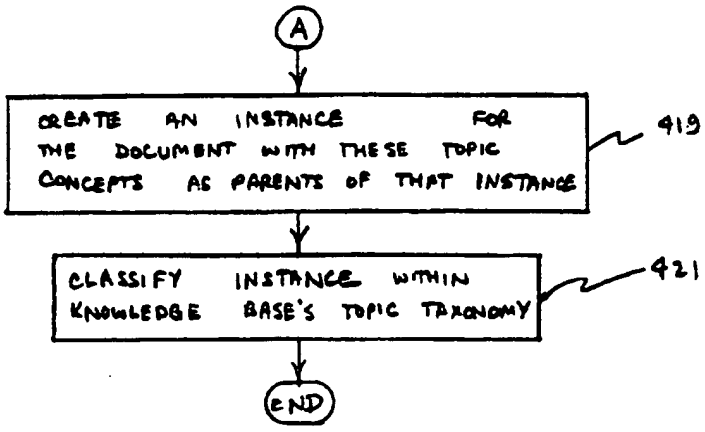

A
CREATE AN INSTANCE FOR THE DOCUMENT WITH THESE TOPIC CONCEPTS AS PARENTS OF THAT INSTANCE
419
CLASSIFY INSTANCE WITHIN KNOWLEDGE BASE'S TOPIC TAXONOMY
421
END

FIG. 4

INTERNATIONAL SEARCH REPORT

Inte nal Application No
PCT/US 97/05782

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 95 12173 A (TELTECH RESOURCE NETWORK CORP) 4 May 1995 see page 5, line 1 - page 13, line 1 ---	1-3,6, 8-17
A	BALDAZO R: "NAVIGATING WITH A WEB COMPASS" BYTE, vol. 21, no. 3, 1 March 1996, page 97/98 XP000600179 see the whole document --- -/--	1-17

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 August 1997	18.08.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Fournier, C

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int nal Application No
PCT/US 97/05782

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GINSBERG A: "A UNIFIED APPROACH TO AUTOMATIC INDEXING AND INFORMATION RETRIEVAL" IEEE EXPERT, vol. 8, no. 5, 1 October 1993, pages 46-56, XP000413472 see page 46, left-hand column, line 1 - page 48, left-hand column, line 14 ---	1,11, 14-16
X	JACOBS P S ET AL: "SCISOR: EXTRACTING INFORMATION FROM ON-LINE NEWS" COMMUNICATIONS OF THE ASSOCIATION FOR COMPUTING MACHINERY, vol. 33, no. 11, 1 November 1990, pages 88-97, XP000173090 see page 89, left-hand column, line 15 - page 92, left-hand column, line 18 ---	1,11, 14-16
A	WO 90 08360 A (TELEBASE SYSTEMS INC) 26 July 1990 see page 5, line 13 - page 8, line 1 ---	2,3,6,12
A	PATENT ABSTRACTS OF JAPAN vol. 018, no. 417 (P-1781), 4 August 1994 & JP 06 124308 A (FUJITSU LTD), 6 May 1994, see abstract -----	1,3,11, 16

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte nal Application No
PCT/US 97/05782

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9512173 A	04-05-95	CA 2175187 A	04-05-95
		US 5634051 A	27-05-97
WO 9008360 A	26-07-90	AU 5024990 A	13-08-90